Optimal Monte Carlo Updating

Lode Pollet,* Stefan M.A. Rombouts, Kris Van Houcke, and Kris Heyde

Vakgroep subatomaire en stralingsfysica,
Proeftuinstraat 86,
Universiteit Gent, Belgium
(Dated: February 2, 2008)

Based on Peskun's theorem it is shown that optimal transition matrices in Markov chain Monte Carlo should have zero diagonal elements except for the diagonal element corresponding to the largest weight. We will compare the statistical efficiency of this sampler to existing algorithms, such as heat-bath updating and the Metropolis algorithm. We provide numerical results for the Potts model as an application in classical physics. As an application in quantum physics we consider the spin $3/2 XY$ model and the Bose-Hubbard model which have been simulated by the directed loop algorithm in the stochastic series expansion framework.

PACS numbers: 02.70.Tt, 05.50.+q, 82.20.Wt

Monte Carlo methods are nowadays used in almost every branch of science, offering exact results in a statistical sense or providing answers where other methods fail. Already in statistical physics alone, Monte Carlo methods have been applied to a variety of models [1]. For many applications, good algorithms have been devised and there exist now solutions to many problems that were initially untractable. A well known example is the critical slowing down in the neighborhood of critical points that has been overcome in both classical (by cluster algorithms [2] [3]) and quantum Monte Carlo (by the loop algorithms [4]). The need for better performing algorithms is clear: efficient algorithms lead to more accurate results at the same computational cost. Yet little that goes beyond common sense reasoning is known about why an algorithm is efficient or not, and within a chosen algorithm there is often additional freedom.

We address the question of the efficiency of Markov chain Monte Carlo (MCMC) algorithms in terms of smaller error bars. We first touch upon the needed terminology as it is usually [5] understood in statistical physics from a practitioner's viewpoint. We show how optimal sampling enters into this discussion and comment on its implementation. Finally, we compare it with standard updating mechanisms for the Potts model and for the directed loop algorithm [6, 7] in the stochastic series expansion [8].

In MCMC a transition kernel (matrix) T is set up and we will assume that we know the discrete weights $W_1, \ldots, W_n$ (finite, computable set) of the invariant probability distribution W. If the following two conditions hold

   (i) normalization of probability, $\sum_j T_{ij} = 1, \forall i$

   (ii) reversibility (detailed balance), $W_i T_{ij} = W_j T_{ji}$,

*Electronic address: Lode.Pollet@UGent.be

and the chain can connect any two states in a certain finite number of steps, then the Markov chain will converge to the invariant probability distribution (which will be W). The stochastic matrix T has as largest eigenvalue 1, while the other eigenvalues are sorted by $-1 < \lambda_j < 1, j = 2, \ldots n$. Strictly speaking, the condition (ii) is a too strong [9] [10] condition to assure convergence of the Markov chain towards the invariant distribution W, it suffices $\sum_i W_i T_{ij} \sim W_j$, but the reversibility condition is widely used in practical applications.

The Markov process correlates the measurements of the observables Q in consecutive steps. The variance σ_Q^2 on these correlated measurements is not equal to the variance $\sigma_{0,Q}^2$ obtained from uncorrelated measurements. Instead, $\sigma_Q^2 = 2\tau_{int,Q}\sigma_{0,Q}^2$, in which we have introduced the integrated autocorrelation time [5] [11],

$$\tau_{int,Q} = \frac{1}{2} + \sum_{t=1}^{\infty} A_Q(x^{(t)}). \tag{1}$$

Stationary samples $x^{(t)}$ at the Monte Carlo times t are obtained from the sampler while the normalized autocorrelation function $A_Q(x^{(t)})$ for the observable Q is given by

$$A_Q(x^{(t)}) = \frac{\langle Q(x^{(i+t)})Q(x^{(i)})\rangle - \langle Q(x^{(i)})\rangle^2}{\langle Q^2(x^{(i)})\rangle - \langle Q(x^{(i)})\rangle^2}, \tag{2}$$

in which the ensemble average $\langle \ldots \rangle$ is taken over i. We can now make a connection with the second largest eigenvalue by

$$\sup_Q \tau_{int_Q} = \frac{1 + \lambda_2}{2(1 - \lambda_2)}. \tag{3}$$

The following discussion focuses on the eigenvalues $\lambda_2, \lambda_3, \ldots$ to obtain a lower asymptotic variance for an observable Q,

$$v(Q, T) = \lim_{n \to \infty} \frac{1}{n}\text{var}\left[\sum_{k=1}^{n} Q(x^{(k)})\right]. \tag{4}$$

A different question concerns the convergence [12] of a probability distribution towards the invariant probability distribution. It is dominated by the second largest eigenvalue in absolute value of the stochastic matrix, which can be different from λ_2 for non-positive operators, and would determine the required number of thermalization or burn-in steps. Note that non-reversible transition kernels can converge faster [13].

The stochastic matrix has the dimension of the Hilbert space, and all algorithms consist of two different operations in every Monte Carlo step: a limitation on the configurations that can be reached and secondly the acceptance or the rejection of the transition to one of them. For instance, heat-bath updating (also called the Gibbs sampler [14]) in the Ising model with dimension $L \times L$ can in one step reach only $L^2 + 1$ different configurations from the current configuration which has weight W_1. Among the L^2 new ones, it picks one at random and the transition to this trial configuration with weight W_2 will be accepted according to $\frac{W_2}{W_1 + W_2}$, otherwise it remains in the current configuration. Note that all eigenvalues of the Gibbs sampler are positive [15].

We will now focus on this second step of the update. For the Ising model, there are only two different configurations that play a role. How should we choose the transition matrix so that the asymptotic variance is smallest? A hint is given by the Peskun theorem [16], stating that if T^A and T^B both satisfy the conditions (i) and (ii) and if all off-diagonal elements of T^B are larger or equal than the corresponding elements of T^A, then T^B will lead to a smaller asymptotic variance for all observables than T^A, or equivalently, $\lambda_2^A > \lambda_2^B$. It then follows that the Metropolis-Hastings algorithm [17] is by construction the most effective sampler for the Ising model with random single site updates (and not the Gibbs sampler). For all possible stochastic matrices with dimension $n = 2$ the Metropolis transition matrix [18] is given by

$$T_{ij}^{Met} = \begin{bmatrix} 0 & 1 \\ \frac{W_1}{W_2} & 1 - \frac{W_1}{W_2} \end{bmatrix}. \qquad (5)$$

Here we have ordered the weights in ascending order. This non-standard way of writing Metropolis updating shows however the key ingredients of its efficiency, namely that the chance of staying in the current configuration should be minimized and secondly that the second largest eigenvalue is $\lambda_2 = -T_{21}^{Met} = -\frac{W_1}{W_2}$.

Peskun's theorem implies an ordering of the weights. So let (for the remaining of the paper) $\pi_1 \leq \pi_2 \leq \ldots \leq \pi_n$ be the normalized weights in ascending order, $\pi_i = \frac{W_i}{\sum_j W_j}$. Peskun's theorem tells us that we can always improve a transition matrix T by 'Metropolizing' it, $T_{ij}' = \frac{T_{ij}}{\sum_{j \neq i} T_{ij}}, \forall j \neq i$. Applying this idea to heat-bath updates, the following Metropolized Gibbs sampler(MG)

is obtained,

$$T_{ij}^{MG} = \begin{bmatrix} 0 & \frac{\pi_2}{1-\pi_1} & \frac{\pi_3}{1-\pi_1} & \cdots & \frac{\pi_n}{1-\pi_1} \\ \frac{\pi_1}{1-\pi_1} & 1 - \cdots & \frac{\pi_3}{1-\pi_2} & \cdots & \frac{\pi_n}{1-\pi_2} \\ \frac{\pi_1}{1-\pi_1} & \frac{\pi_2}{1-\pi_2} & 1 - \cdots & \cdots & \frac{\pi_n}{1-\pi_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\pi_1}{1-\pi_1} & \frac{\pi_2}{1-\pi_2} & \frac{\pi_3}{1-\pi_3} & \cdots & 1 - \cdots \end{bmatrix}, \qquad (6)$$

or $T_{ij}^{MG} = \min(\frac{\pi_j}{1-\pi_i}, \frac{\pi_j}{1-\pi_j})$. Liu [19] has applied this idea to the independence Gibbs sampler, and obtained a complete eigenanalysis for the resulting stochastic matrix.

It is possible to repeat this Metropolizing procedure until all but one of the diagonal elements are zero. So, optimal transition matrices must have $T_{ii} = 0, i \neq n$. Indeed, Frigessi *et al.* [20] have shown that the optimal transition matrix is of the form,

$$T_{ij}^{Opt} = \begin{bmatrix} 0 & \frac{W_2}{W_1}y_1 & \frac{W_3}{W_1}y_1 & \cdots & \frac{W_n}{W_1}y_1 \\ y_1 & 0 & \frac{W_3}{W_2}y_2 & \cdots & \frac{W_n}{W_2}y_2 \\ y_1 & y_2 & 0 & \cdots & \frac{W_n}{W_3}y_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_1 & y_2 & y_3 & \cdots & 1 - y_1 - y_2 - \cdots \end{bmatrix}, \qquad (7)$$

with $y_1 = \frac{\pi_1}{1-\pi_1}, y_2 = (1 - y_1)\frac{\pi_2}{1-\pi_1-\pi_2}, \ldots$. The eigenvalues are given by $1, \lambda_2 = -y_1$ (the same as in Eq.(6)), $\lambda_3 = -y_2, \ldots$. They are all negative and appear in an ordered way. This λ_2 has the lowest value that possibly can be obtained with respect to the probability distribution W, and with λ_2 determined, λ_3 is then the smallest possible third largest eigenvalue, etc. Note that a rescaling is at work here, the entries for the second row $T_{2j}, j = 2, \ldots, n$ are analogous to the first row apart from the rescaling $1 \rightarrow (1 - y_1)$. Eq.(7) represents an optimal transition matrix over the entire Hilbert space, however, many situations of practical interest need to sample stochastic subprocesses. Within these, optimal sampling can only be achieved when all but one of the diagonal elements are zero. When Eq.(7) is applied to a stochastic subprocess, we will call it the locally optimal algorithm.

Potts model. As a first application, we consider the $q = 4$ Potts model [1] in two dimensions. We are interested in the dynamics of the Monte Carlo process. Therefore we consider a small lattice with single-spin updates only, and we do not want to use cluster updates [2] [3] here. So we will randomly select a spin, after which we have to make a choice between the four possible orientations that this spin can take. Although the random selection of a site and the single spin update both seriously violate the structure of the optimal stochastic matrix Eq.(7), the Peskun theorem still holds, meaning that the matrix Eq.(7) (applied to the stochastic subprocess) still leads to a more effective sampling than heat-bath updating, $T_{ij} = \frac{W_j}{\sum_k W_k} = \pi_j, \forall i, j$. This of course cannot cure the fact that the updating of a single spin still leads
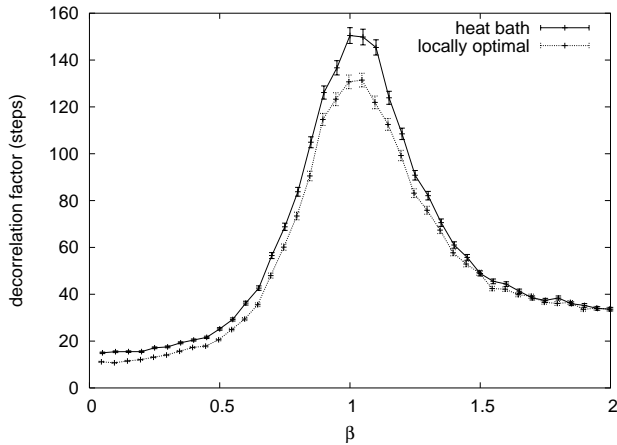
FIG. 1: The decorrelation factor for the energy as a function of temperature β at constant interaction strength for the $q = 4$ Potts model on a $L = 4 \times 4$ matrix is shown for heat-bath updates and for the locally optimal ones. Simulations consisted of 4000 chains of one million steps each.

to the divergence of the autocorrelation time when the temperature reaches its critical value β_c.

The decorrelation factor [21] $\sigma_Q^2/\sigma_{0,Q}^2$ is defined as the ratio between the error bars obtained from a Monte-Carlo simulation with correlations between successive samples and the error bars one would obtain from the same number of independent but identically distributed samples. In the limit of a large number of samples, the decorrelation factor becomes equal to twice the integrated autocorrelation time (Eq.(1)). The decorrelation factor can accurately be estimated by running a large number of independent Markov chains. In Fig. 1 we see that the decorrelation factor is much smaller for the locally optimal algorithm than for the heat-bath algorithm.

A small lattice of $L = 4 \times 4$ has been chosen since the decorrelation factor scales with system size for single-site updates such that the difference between the two algorithms will be larger in absolute terms when using a small lattice. We clearly see a significant difference between the locally optimal and the heat-bath transition probabilities. The Metropolized Gibbs sampler T^{MG} (Eq.(6)) has the same λ_2 as the optimal one (Eq.(7)), and the integrated autocorrelation times differ only slightly for the 4×4 Potts model. Note that it is much easier to implement T^{MG} than T^{Opt}, and if in practice the stochastic matrices cannot be computed in advance and need to be recomputed at each step, it is recommended to use T^{MG}.

Directed Loops. The same reasoning also applies to quantum Monte Carlo. In the stochastic series expansion method [8] a Taylor expansion is applied to the partition function $Z = Tr \exp(-\beta H)$, yielding

$$
\begin{aligned}
Z = \sum_{m=0}^{\infty} \frac{\beta^m}{m!} \sum_{\{i_1,\dots,i_m\}} \sum_{\{b_1,\dots,b_m\}} \langle i_1| - H_{b_1}|i_2\rangle\langle i_2| - H_{b_2}|i_3\rangle \\
\cdots \langle i_{m-1}| - H_{b_{m-1}}|i_m\rangle\langle i_m| - H_{b_m}|i_1\rangle,
\end{aligned} \tag{8}
$$

where the Hamiltonian H is decomposed in a set of bond operators, $H = \sum_b H_b$, and complete sets $|i_k\rangle, k = 1,\dots,n$ have been inserted. In the first step, a diagonal update is performed in which the expansion order m can change. The second, off-diagonal update mimics the idea of loop-type [4] and worm [22] updates and can best be explained using a graphical interpretation. Every matrix element in Eq.(8) is called a vertex, at which the two sites of the interaction and an imaginary time are assigned. Every vertex has four legs: two incoming and two outgoing legs per site, corresponding to the particles created and annihilated by H_b. The legs of the vertices are connected by segments corresponding to the occupied sites. A worm is created in a arbitrary point in space-time by inserting a creation and annihilation operator on a segment. One of these operators is chosen to be the worm head and is mobile, while the other one is the worm tail and remains immobile. The worm head moves through configuration space and can change the type of the vertices, for instance from a diagonal to an off-diagonal vertex. The worm movement stops when the worm head bites into its own tail. The entire movement of the worm can then be regarded as a single loop update.

Once the worm has been created, its movement is completely determined by how it passes through and modifies the vertices. Suppose the worm head reaches a vertex at the leg left under (entrance leg), as in Fig. 2. The local configuration can then change according to the four processes bounce, straight, jump and turn, each with their proper weight $W_i, i = 1,\dots,n$. For models with number conservation, it is possible that one or more of the four processes cannot occur, so n can in principle be two, three or four. The bounce process is always possible but since it does not change the current vertex, it can be regarded as a waste of computer time. The worm head has to choose between one of the n processes, modifies hereby the current vertex and goes consequently to the next vertex along the segment that connects the current exit leg and the next entrance leg. The probability matrix T_{ij} defines the transition from the entrance leg to the exit legs and hence completely determines the worm movement. We will now discuss several choices for this probability matrix T_{ij}.

Originally, the heat-bath updates [8], $T_{ij} = \frac{W_i}{\sum_j W_j}$, (solution A) were proposed. Secondly, other choices are perfectly possible: Syljuåsen and Sandvik propose *directed loops* [6], where the worm head has a preferred direction at the vertices in order to be more efficient than solution A. The rule of thumb is that the frequency of the bounce processes should be as low as possible. This inspired the authors of Ref. [23] to numerically
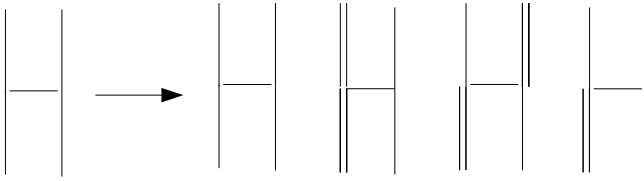
FIG. 2: The four possible states that can arise when the worm enters the leftmost vertex at the leg left under in a number occupation basis (can also be a spin state) and for a system with particle number conservation. A single line means that the leg is singly occupied, a double line means double occupancy and a dashed line denotes that the leg is not occupied. The four processes on the right correspond to bounce, straight, jump and turn, from left to right.

minimize the trace of the probability matrix T_{ij} with respect to detailed balance, $W_i T_{ij} = W_j T_{ji}$ (hereafter called solution B or the minimal bounce solution). They used a linear programming technique [24] for this purpose. Equivalently, one can say that this amounts to minimizing the sum of the eigenvalues of the transition matrix T_{ij}, $\min(\lambda_2 + \lambda_3 + \lambda_4)$. Thirdly, we propose to use the locally optimal probability matrix Eq.(7) as transition matrix (solution C) for the scattering of the worm at a vertex.

Spin $3/2XY$ model. In Ref. [23] the directed loop algorithm was studied for the one dimensional spin $3/2XY$ model in an external magnetic field h,

$$ H = J \sum_{\langle i,j \rangle} \frac{1}{2} (S_i^+ S_j^- + S_i^- S_j^+) - h \sum_i S_i^z, \qquad (9) $$

where the first sum extends over nearest neighbors and J is an exchange interaction term. It appeared that solution B always gave shorter autocorrelation times than solution A, as can also be seen in Fig. 3 and in Fig. 4, where the integrated autocorrelation times for the uniform magnetization and the energy are plotted.

The authors of Ref. [23] also proposed to break detailed balance to $f_i W_i T_{ij} = f_j W_j T_{ji}$, with f_i an extra degree of freedom at leg i. Using a linear programming [24] technique they were able to further minimize the trace of the probability matrix. They found that this algorithm gave the shortest autocorrelation times for most values of the magnetic field, except around $h \approx 0.8$ where it behaved worse than solution B and, unexpectedly, even worse than solution A. They deduced that alternative principles than minimizing bounces might exist. Furthermore, this algorithm needs to be used carefully, since it modifies condition (ii) with respect to the invariant distribution for the Green's function [25] $\langle a_i^\dagger(0) a_j(t) \rangle$, although it is correctly weighted with respect to the invariant distribution for diagonal observables such as the energy and the magnetization. Note that it is also possible to apply Eq.(7) to the modified weights $W_i' = f_i W_i$ in order to obtain the locally optimal algorithm in case one is not
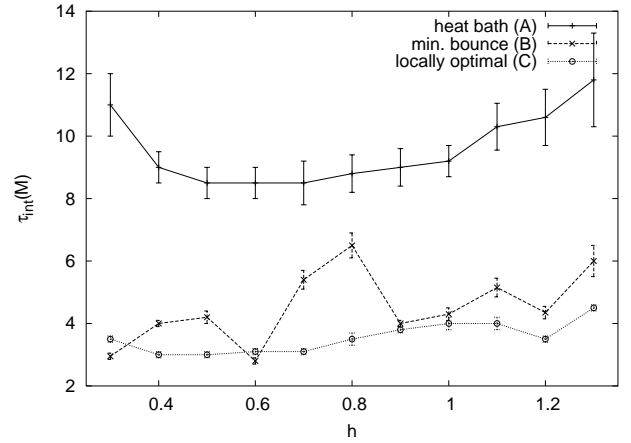

FIG. 3: Integrated autocorrelation times for the magnetization, $\tau_{int}(M)$, as a function of magnetic field h for a spin $3/2XY$ model as in Ref. [23], lattice size $L = 64$, inverse temperature $\beta = 64$. The integrated autocorrelation times are made loop size independent (normalized to two worms per update) so that the heat bath (solution A), minimal bounce (solution B) and locally optimal (solution C) algorithms can directly be compared. The precise definition of these algorithms is explained in the text.

interested in the Green's function. Therefore we will not further compare this proposal with solutions A, B and C.

We also addressed the spin $3/2XY$ model with solution C and the results for the integrated autocorrelation times for the magnetization and the energy are also presented in Fig. 3 and in Fig. 4. We find a substantial improvement over solution A. Solution C is also much better than solution B for magnetic fields around $h \approx 0.8$, while for other magnetic fields the difference is smaller. This indeed shows that minimizing bounces is not optimal for the spin $3/2XY$ model. Note that increasing the system size does not significantly change the ratio of the correlation times of solutions B and C.

Bose-Hubbard model. We also present results for the Bose-Hubbard model [26] in one dimension (units are as in Ref. [27]),

$$ H = -t \sum_{\langle i,j \rangle} b_i^\dagger b_j + \frac{1}{2} U \sum_i n_i (n_i - 1) - \sum_i \mu n_i. \quad (10) $$

The first term represents hopping with strength t of the bosons between nearest neighbors, the chemical potential is denoted by μ and the second term takes on-site repulsion with strength U into account. In the Bose-Hubbard model, the diagonal weights are relatively much larger compared to the non-diagonal weights than in spin systems. Again, as can be seen in Fig. 5, the heat-bath updates (solution A) are outperformed, but for large on-site repulsion U the minimal bounce solution (solution B) is superior to the locally optimal solution (solution C).
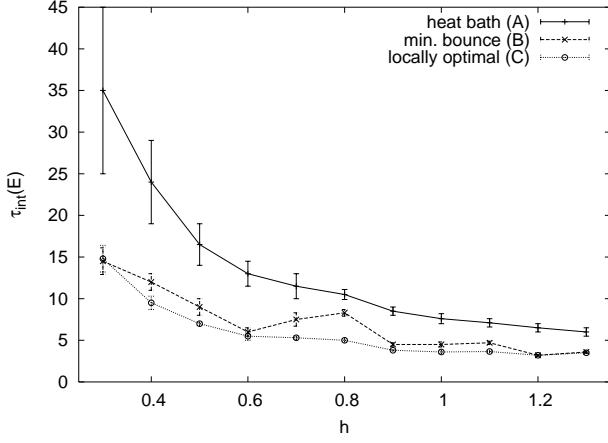
FIG. 4: Analogous to Fig. 3 but now for the integrated autocorrelation time of the energy, $\tau_{int}(E)$, as a function of the magnetic field h.
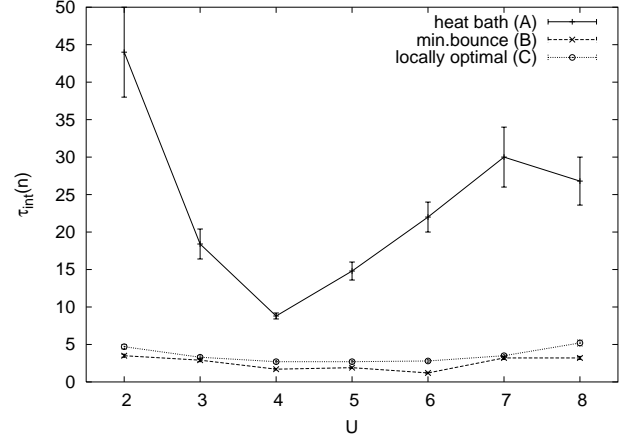

FIG. 5: Integrated autocorrelation time for the density, $\tau_{int}(n)$, as a function of on-site repulsion U for a one dimensional Bose-Hubbard model with $\mu = 5, t = 1$, lattice size $L = 64$ and inverse temperature $\beta = L$. Four loops were constructed in every update for the 'min. bounce (B)' and 'locally optimal (C)' algorithms, while for 'heat bath (A)' we constructed 16 loops and multiplied the results by four afterwards. Particle number cut-off is lowered at $U = 3$ and $U = 8$. The Mott phase sets in for $U > 9$.

As in the Potts model, we are guaranteed that solution C is more efficient than solution A, but the Peskun theorem does not claim that solution C is superior to solution B, because choosing the lowest (local) λ_2 will not necessarily correspond to the lowest integrated autocorrelation time, τ_{int} (Eq.(1). Specifically, in case $n = 2$ both solutions reduce to Metropolis updating Eq.(5). In case $n = 3$ solution B is of the form

$$T_{ij}^B = \begin{bmatrix} 0 & a & 1-a \\ \frac{\pi_1}{\pi_2}a & 0 & 1-\frac{\pi_1}{\pi_2}a \\ \frac{\pi_1}{\pi_3}(1-a) & \frac{\pi_2}{\pi_3}-\frac{\pi_1}{\pi_3}a & \frac{2\pi_3-1+2\pi_1 a}{\pi_3} \end{bmatrix}. \quad (11)$$

The linear programming technique [24] applied in solution B will try to make $a = \frac{1-2\pi_3}{2\pi_1}$ if $\pi_3 < \frac{1}{2}$, otherwise it will take $a = 0$. It will be the system parameters that determine whether T^B or T^C (Eq.(7) for $n = 3$) performs better. Due to its structure it is also not possible to improve T_B by Metropolizing it.

Also in case $n = 4$ both solutions B and C put all diagonal elements in the (local) stochastic matrix zero, except for the diagonal element corresponding to the largest weight. We know from the Peskun theorem that this leads to an efficient sampling. As a counterexample, the Metropolized Gibbs sampler Eq.(6) has only one zero on the diagonal of the transition matrix and is systematically outperformed by both solution B and C.

The class of locally optimal stochastic matrices can be parameterized as

$$T_{ij} = \begin{bmatrix} 0 & a & b & x \\ \frac{\pi_1}{\pi_2}a & 0 & c & y \\ \frac{\pi_1}{\pi_3}b & \frac{\pi_2}{\pi_3}c & 0 & z \\ \frac{\pi_1}{\pi_4}x & \frac{\pi_2}{\pi_4}y & \frac{\pi_2}{\pi_4}z & 1-\frac{\pi_1 x+\pi_2 y+\pi_3 z}{\pi_4} \end{bmatrix}, \quad (12)$$

with $x = 1-a-b, y = 1-\frac{\pi_1}{\pi_2}a-c, z = 1-\frac{\pi_1}{\pi_3}b-\frac{\pi_2}{\pi_2}c$ and the three free parameters a, b and c. Solution B will now try to minimize T_{44}, or minimize a, b and c, under a number of constraints such as $a + b \le 1$. This can be suboptimal

however, since the minimum will be found when one or more of the constraints are exactly met [24]. Suppose the minimum is found for $a + b = 1$, as it happens in the spin $3/2 XY$ model for magnetic fields $h \approx 0.8$. Then the scattering from the least probable state to the most probable state is zero, and this clearly cannot be optimal and explains why solution C is better in Fig. 3.

The overall conclusion is that the integrated autocorrelation times for both solutions B and C will be of the same order and roughly optimal. The important principle is that the diagonal elements corresponding to the lowest $n - 1$ weights should be zero. Some arbitrariness is still retained, but it does not seem possible to define how the remaining freedom should be chosen independently of the weights that occur in the updating process. When the diagonal and non-diagonal weights are of the same order, solution C is better, while for large diagonal weights solution B gives the lowest integrated autocorrelation times. Furthermore, we have a good argument why these solutions lead to a locally optimal sampling, based on the Peskun theorem.

Conclusion. We have shown that the optimal transition kernel for Markov chain Monte Carlo should have a zero diagonal, except for the diagonal element corresponding to the largest weight (which can be large). We have presented results for the Potts model with random single spin updates and for quantum spin chains and the Bose-Hubbard model with the directed loop algorithms. They confirm the theoretical reasoning. Our results suggest a practical way to improve existing Monte Carlo methods. This could lead to significant gains in

efficiency in both classical and quantum Monte Carlo. One could consider applications in the research fields of flat histogram methods [28], the loop algorithm [4] [5], the worm algorithm [22] and the fast updates in auxiliary field quantum Monte Carlo [29] [30] and shell model Monte Carlo [31]. Furthermore, our results could shed a new light on the correlations in Glauber dynamics [32].

[1] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics*, Oxford University Press, Oxford (1999).

[2] R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[3] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[4] H. G. Evertz, G. Lana and M. Marcu, Phys. Rev. Lett. **70**, 875 (1993).

[5] H. G. Evertz, Adv. Phys. **52** 1 (2003).

[6] O. F. Syljuåsen and A. W. Sandvik, Phys. Rev. E **66**, 046701 (2002).

[7] O. F. Syljuåsen, Phys. Rev. E **67**, 046701 (2003).

[8] A. W. Sandvik, Phys. Rev. B **59**, R14157 (1999).

[9] L. Tierney, Annals of Statistics, **22**, 1701 (1994).

[10] V. I. Manousiouthakis and M. W. Deem, J. Chem. Phys. **110** (6), 2753-2767 (1999).

[11] W. Janke, *Statistical Analysis of Simulations: Data Correlations and Error Estimation* in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, eds. J. Grotendorst, D. Marx and A. Muramatsu, John von Neumann Institute for Computing, NIC-series vol. 10, Jülich (2001).

[12] A. Mira, *PhD-thesis*, U. of Minnesota (1998).

[13] P. Diaconis, S. Holmes and R. M. Neal, Ann. of Appl. Prob. **10**, 726 (2000).

[14] S. Geman and D. Geman, IEEE Transactions on Pattern Analysis and Machine Intelligence, **6**, 721 (1984).

[15] J. S. Liu, W. H. Wong and A. Kong, J. Roy. Statist. Soc. Ser. B, **57**, 157 (1995).

[16] P. H. Peskun, Biometrika **60**, 607 (1973).

[17] W. K. Hastings, Biometrika **57**, 97 (1970).

[18] N. Metropolis, A. W. Rosenbluth, M. N. Metropolis, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[19] J. S. Liu, Statistics and Computing **6**, 113 (1996); J. S. Liu, Biometrika **83**, 681 (1996).

[20] A. Frigessi, C. R. Hwang and L. Younes, Ann. of Appl. Prob. **2**, 610 (1992).

[21] S. M. A. Rombouts, *PhD-thesis*, Gent U. (1997) (see http://allserv.ugent.be/~srombout/thesis.ps.gz).

[22] N.V. Prokof'ev, B.V. Svistunov and I.S. Tupitsyn. JETP Lett. **64**, 911 (1996).

[23] F. Alet, S. Wessel and M. Troyer, cond-mat/0308495 (2003).

[24] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge (1992).

[25] A. Dorneich and M. Troyer, Phys. Rev. E **64**, 066701 (2001); N. Kawashima and K. Harada, J. Phys. Soc. Jpn. **73**, 1379 (2004).

[26] M. V. Zyubin and V. A. Kashurnikov, Phys. Rev. E **69**, 036701 (2004).

[27] L. Pollet, S. Rombouts, K. Heyde, and J. Dukelsky, Phys. Rev. A **69**, 043601 (2004).

[28] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[29] F. F. Assaad, *Quantum Monte Carlo Methods on Lattices: The Determinantal Approach* in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, eds. J. Grotendorst, D. Marx and A. Muramatsu, John von Neumann Institute for Computing, NIC-series vol. 10, Jülich (2001) and references therein.

[30] S. M. A. Rombouts, K. Heyde and N. Jachowicz, Phys. Rev. Lett. **82**, 4155 (1999).

[31] S. E. Koonin, D. J. Dean and K. Langanke, Phys. Rep. **278**, 1 (1997).

[32] R. J. Glauber, J. Math. Phys **4**, 294 (1963).